

AI + Big Data 下统计辉煌时代来临

■ 谢邦昌 叶玲珑

在 AI + Big Data (人工智能 + 大数据) 迅猛发展的背景下, 统计不应“抱残守缺”, 纠结于显著性的困境, 而是应该与时俱进, 开启未来几年甚至十几年“翻转”发挥的辉煌时代。在快速运算大量储存的时代里, 统计的研究服务教学也应该因此变革, 以数据分析为依据迈向 AI 时代的来临。

人工智能在未来 10 年会有三个阶段, 第一阶段, 弱人工智能: 擅长单一方面, 像是辨别语音、辨别影像, AlphaGo → 医疗、Amazon → 物流、IBM → Watson (认知计算), 目前的 AI 还是处于弱人工智能, 5G 在未来 5 年成熟之后, 就会进入强人工智能时代; 第二阶段, 强人工智能: 相当于人类级别的人工智能, 在未来 10 年内, 量子计算机发展成熟后, 就会进入超人工智能时代; 第三阶段, 超人工智能: 几乎所有领域具备比单个人类更强大的智慧。在接下来的 10 年内, 必须要面对 AI 的卓越成长, 并且要辅助 HI (人类智能), 让 AI 辅助人类的生活, 使人类的生活能够更进步。

统计上, 简单易用的 P 值常用于衡量显著性。但是, 在 P 值的应用场景中, 实际分析的情况往往难以满足严格的假设条件, 再加上使用错误, 以及对具体问题解读的偏差, 容易导致无法重复或矛盾的结论。尽管 P 值的可靠性备受争议, 其在研究报告中的使用却越来越流行。

为此, 今年 3 月有三位统计学家在美国《自然》杂志上发表了一封公开信, 呼吁研究人员放弃统计显著性, 停止采用传统二分法划分 P 值作为分析结果是否支持某一研究假设的判断标准。他们的呼吁在短短一周内得到了 800 多名研究人员的支持, 再次引发了反对统计显著性的讨论热潮。

为了跳出显著性的困境, 三位统计学家建议, 在对科学问题进行判断时, 在统计上不应该一刀切, 必

须要学会承认和接受不确定, 以谦虚的方式解读兼容性区间。与过去显著性困境不同的是, 如今正蒸蒸日上的 AI + Big Data 让我们有机会利用统计去探索科学研究及现实应用中更多的不确定性与兼容性。

诺贝尔经济学奖获得者托马斯·萨金特去年在我国公开演讲时指出: “人工智能其实就是统计学, 只不过用了一个很华丽的辞藻, 其实就是统计学。”托马斯的论断一度引发学界与业界的热议。尽管这个论断较为片面, 但从一定程度上强调了统计在 AI 及 Big Data 时代的重要地位。

AI 的本质是基于 Big Data 的学习和推理, 而统计不仅是解决学习算法和推理模型的重要基础之一, 统计观念更是对整个 AI 的实现过程——包括 Big Data 中的数据获取、数据融合、数据清洗、目标定义以及最终结果的判断与呈现都有着巨大的影响。基于 AI 及 Big Data, 统计不再仅仅靠算法或模型驱动, 还应该具备由问题驱动的数据思维, 以应对实际数据的不确定性和兼容性。

当然, 诚如反对托马斯论断的片面性, AI 及 Big Data 也不仅仅是统计学, 其以开放的态度, 在不同方向上不断吸收和借鉴数学、运筹学、计算机科技、神经科学和心理学等众多学科, 以在充满不确定性和考验兼容性的环境下, 实现真正的价值。

随着 AI 及 Big Data 进入发展的快车道, 统计同样应该拥抱不确定性与兼容性, 迈向未来几年甚至十几年的辉煌时代。AI 及 Big Data 时代绝对少不了统计人的参与, 统计学更是此时代的重要领域。“天下武功, 无坚不摧, 唯快不破”是大家都知的速度为王, 而今更是数据为王的时代, 身为统计人责无旁贷, 让我们一起迎接统计辉煌时代的来临, 迈向 AI + HI (人工智能 + 人类智能) 的新纪元。❏

作者单位: 台北医学大学管理学院
厦门大学公共卫生学院